Human Secreted Carbonic Anhydrase: cDNA Cloning, Nucleotide Sequence, and Hybridization Histochemistry[†,‡]

Peter Aldred,[§] Ping Fu, Graham Barrett,[∥] Jennifer D. Penschow, R. Douglas Wright, John P. Coghlan, and Ross T. Fernley*

The Howard Florey Institute of Experimental Physiology and Medicine, Parkville, Victoria 3052, Australia

Received July 17, 1990; Revised Manuscript Received September 19, 1990

ABSTRACT: Complementary DNA clones coding for the human secreted carbonic anhydrase isozyme (CA VI) have been isolated and their nucleotide sequences determined. These clones identify a 1.45-kb mRNA that is present in high levels in parotid submandibular salivary glands but absent in other tissues such as the sublingual gland, kidney, liver, and prostate gland. Hybridization histochemistry of human salivary glands shows mRNA for CA VI located in the acinar cells of these glands. The cDNA clones encode a protein of 308 amino acids that includes a 17 amino acid leader sequence typical of secreted proteins. The mature protein has 291 amino acids compared to 259 or 260 for the cytoplasmic isozymes, with most of the extra amino acids present as a carboxyl terminal extension. In comparison, sheep CA VI has a 45 amino acid extension [Fernley, R. T., Wright, R. D., & Coghlan, J. P. (1988b) *Biochemistry 27*, 2815]. Overall the human CA VI protein has a sequence identity of 35% with human CA II, while residues involved in the active site of the enzymes have been conserved. The human and sheep secreted carbonic anhydrases have a sequence identity of 72%. This includes the two cysteine residues that are known to be involved in an intramolecular disulfide bond in the sheep CA VI. The enzyme is known to be glycosylated and three potential N-glycosylation sites (Asn–X–Thr/Ser) have been identified. Two of these are known to be glycosylated in sheep CA VI. Southern analysis of human DNA indicates that there is only one gene coding for CA VI.

Carbonic anhydrase (EC 4.2.1.1, carbonate dehydratase) catalyses the reversible hydration of carbon dioxide. The enzyme is distributed widely in different species and different tissues. In mammals at least six isozymes are known to occur. These are CA I,[1] II, and III, which are the soluble cytoplasmic isozymes (Venta et al., 1987), CA IV, which is the membrane-bound form (Wistrand, 1984), CA V, a mitochondrial matrix enzyme (Storey et al., 1984), and CA VI, the secreted isozyme. A gene coding for a seventh carbonic anhydrase, possibly another cytoplasmic isozyme, has been described (Venta et al., 1987). The properties of these carbonic anhydrase isozymes have been reviewed recently (Fernley, 1988; Tashian, 1989; Edwards, 1990). The nucleotide sequence of a cDNA clone encoding an as yet unidentified carbonic anhydrase (named CA Y) has been described recently (Amor-Gueret & Levi-Strauss, 1990). This would appear to represent the mitochondrial isozyme, although this awaits conformation. CA VI has been found in the saliva and parotid glands of a number of mammalian species (Fernley et al., 1979; Feldstein & Silverman, 1984; Murakami & Sly, 1987; Fernley et al., 1989). It has a number of properties that distinguish it from the well-characterized cytoplasmic isozymes. In the native state it occurs as an oligomer (Fernley et al., 1979, 1988a) and has an apparent subunit molecular mass of about 45 000 (Fernley et al., 1984; Feldstein & Silverman, 1984; Murakami & Sly, 1987). These enzymes are glycoproteins and removal of the N-linked carbohydrate reduces the molecular mass to 36 000 (Murakami & Sly 1987; Fernley et al., 1988a). The complete amino acid sequence of sheep CA VI has been determined (Fernley et al., 1988b). It has a total of 307 amino acids compared to 259 or 260 for the cytoplasmic isozymes and has a sequence identity of 33% with sheep carbonic anhydrase II. All residues known to be involved in the active site of CA II have been conserved in CA VI. Sheep CA VI is stabilized by a disulfide bond linking Cys 25 to Cys 207.

We describe here the nucleotide and derived amino acid sequence of human secreted carbonic anhydrase, compare its sequence with the other carbonic anhydrases, and examine the tissue expression of the gene by Northern analysis and by hybridization histochemistry.

Materials and Methods

Materials. Restriction endonucleases, T4 DNA ligase, Taq DNA polymerase, Rous Associated Virus 2 reverse transcriptase, and T4 polynucleotide kinase were purchased from Pharmacia or Amersham Corp. Unlabeled nucleotides were obtained from P-L Biochemicals. Radioisotopes ([α-^{32}P]dCTP and [α-^{32}P]dATP) were obtained from Bresatec Australia and Amersham Corp. AMV reverse transcriptase and DNA polymerase were from Life Sciences U.S.A. and Bresatec, respectively. Sequenase was purchased from US Biochemical Co. Nitrocellulose paper was from Schleicher and Schuell (S & S), and Biotrace RP Nylon membranes were from Gelman Sciences Inc. Oligodeoxynucleotides were synthesized by the phosphoramidite procedure (McBride & Caruthers, 1983), using an Applied Biosystems DNA synthesizer. Materials for polyacrylamide gel electrophoresis were obtained from Bio-Rad Laboratories (Richmond, CA), and OCT freeze-embedding

[†] This work was supported by the National Health and Medical Research Council of Australia.

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05305.

* Address correspondence to this author.

[§] Present address: Group Technical Centre, Carlton and United Breweries, Melbourne, Australia.

[∥] Present address: Division of Vascular Medicine and Artherosclerosis, Brigham and Woman's Hospital, Boston, MA.

[1] Abbreviations: CA, carbonic anhydrase; SSC, 150 mM NaCl and 15 mM sodium citrate; SDS, sodium dodecyl sulfate; PCR, polymerase chain reaction; TE, tris (10mM), EDTA (1mM) (pH 7.5) buffer.

0006-2960/91/0430-569$02.50/0

material was from Lab-Tek Products (Naperville, IL). Other chemicals used were at least of analytical grade.

RNA Isolation and Northern Analysis. RNA was isolated from tissues by the guanidinium thiocyanate extraction procedure, followed by ultracentrifugation through a CsCl cushion (Chirgwin et al., 1979). RNA for Northern blot analysis was separated by electrophoresis through 0.8% agarose/2.2 M formaldehyde gels (Maniatis et al., 1982). The RNA was transferred to nitrocellulose filters by capillary blotting, and the filters were baked in vacuo at 80 °C for 2 h. Filters were prehybridized for 2 h at 42 °C in 20% formamide, 5 × SSC, 0.1% bovine serum albumin, 0.1% ficoll 400, 0.1% poly(vinylpyrrolidone), 0.1 mM ATP, and 100 μg/mL denatured salmon sperm DNA in 50 mM Tris (pH 8.0).

Hybridization was performed at 42 °C overnight in the same buffer with the addition of the labeled probe. The probe used was an oligonucleotide (30-mer) corresponding to nucleotides 70–99 in human CA VI mRNA and this was 5′-end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (Maxam & Gilbert, 1980). Filters were washed four times in 2 × SSC and 0.1% SDS, at 40 °C, and three times in 0.1 × SSC and 0.1% SDS each at 50 °C, air dried, and exposed to Kodak XAR-5 film at −80 °C with an intensifying screen for 7 days. The filter was stripped and then rehybridized with a ^{32}P-labeled actin oligonucleotide to demonstrate adequate transfer of RNA to the filter.

Isolation of CA VI Clones. A human salivary gland cDNA library in λ gt10 (Evans et al., 1986) was screened with the ^{32}P-labeled oligonucleotide corresponding to the amino-terminal sequence of the sheep CA VI enzyme (Fernley et al., 1989). Duplicate nitrocellulose filter-lifts were taken (Benton & Davis, 1977) and prehybridized/hybridized as for the Northern analysis except the buffer contained 20% formamide. The filters were washed to a stringency of 2× SSC and 0.1% SDS at 35 °C.

The cDNA inserts were subcloned either complete or after digestion with four-base cutters into M13 mp8. Random subclones were selected and subjected to nucleotide sequence analysis by the chain termination method (Sanger et al., 1977).

For one region of the molecule where no suitable subclones were obtained, two oligonucleotides (17-mers, one sense and the other antisense), corresponding to a sequence already determined, were used as primers in the sequencing reaction to obtain the missing sequence. Primer 1 corresponded to nucleotides 374–390 while primer 2 corresponded to nucleotides 473–489.

To clone the 3′-end of the molecule, the polymerase chain reaction method was used (Saiki et al., 1988). The $d(T)_{17}$ adaptor and adaptor sequences as described by Frohman et al. (1988) were used in conjunction with a 24 base oligonucleotide corresponding to the human CA VI mRNA sequence (nucleotide 809–832). This sequence had been determined from a previously isolated and sequenced clone. Human submandibular gland poly(A+) RNA (3 μg), prepared by the method of Aviv and Leder (1972), was used in a 40-cycle polymerase chain reaction. The product was electrophoresed on an 0.8% low gelling temperature agarose gel and visualized with ethidium bromide, and a 500-bp band was cut out of the gel. The gel was melted in TE buffer at 70 °C, phenol–chloroform extracted, and ethanol precipitated. The purified DNA was cloned into an M13 mp10 vector. Plasmids containing the CA VI inserts were identified by colony-lift hybridization, using the purified PCR fragment as a probe. Mini-prep plasmid DNA was sequenced with Sequenase following the supplier's recommendations.

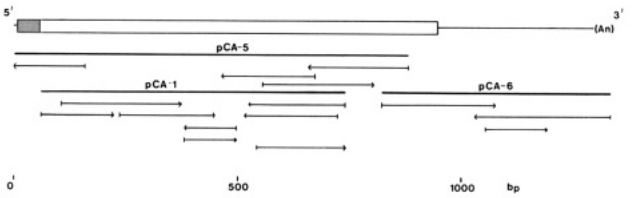

FIGURE 1: Strategy for determining the nucleotide sequence of human CA VI cDNA inserts. The sequence corresponding to the coding region is indicated by the outlined box. The direction and extent of sequence determinations are shown by the horizontal arrows under each clone used; the sites of 5′-end-labeling are indicated by short vertical lines at the end of arrows.

Assembly of sequenced fragments was performed by using the programs of Staden (1977) as modified by Dr. A. Kyne (Walter and Eliza Hall Institute, Melbourne).

Southern Analysis. Human genomic DNA (a gift from Matthew Digby, Howard Florey Institute) was digested with restriction enzymes, electrophoresed in an 0.8% agarose gel, and transferred by using a modification (Clouston et al., 1989) of the Southern procedure (Southern, 1975) in alkali to Biotrace RP Nylon membranes. Hybridization with a ^{32}P-labeled carbonic anhydrase VI gene probe was conducted overnight at 65 °C in 3 × SSC, 0.4% skim milk powder, 1.0% SDS, and 0.5 μg/mL herring sperm DNA. The probe consisted of two oligodeoxyribonucleotide probes corresponding to nucleotides 155–194 (sense strand) and 180–219 (antisense), which overlap by 15 nucleotides leaving single-stranded 5′-extensions of 25 nucleotides. Labeling was carried out by using end-fill reactions with [α-^{32}P]dATP, [α-^{32}P]dCTP, and the Klenow fragment from *Escherichia coli* DNA polymerase I. After hybridization, the filter was washed twice in 2 × SSC and 0.1% SDS and then twice in 0.5 × SSC and 0.1% SDS at 50 °C and exposed for 14 days to X-ray film at −80 °C with an intensifying screen.

Hybridization Histochemistry. Hybridization histochemistry was performed on sections of human parotid gland as described by Penschow et al. (1986). Sections of mouse salivary glands and brain were included as controls. The probe used was a synthetic oligonucleotide (30-mer), corresponding to amino acids 202–211 of the mature protein. The probe was 5′-end-labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase.

Results

Isolation of cDNA Clones Coding for Human CA VI. cDNA clones specific for the human secreted carbonic anhydrase isozyme (CA VI) were isolated from a human salivary gland cDNA library by hybridization with a synthetic oligonucleotide on the basis of the known amino-terminal protein sequence of sheep CA VI. It has been shown previously that this oligonucleotide hybridizes to a specific band in sheep parotid gland mRNA and cross-hybridizes to a similar mRNA species in human salivary gland mRNA by Northern analysis (Fernley et al., 1989). This probe turned out to have a 68% identity with the determined nucleotide sequence for human CA VI with mismatches clustered at the 5′-end of the sequence. These clones encoded human CA VI from the initiator Met to amino acid 270. Following the codon for residue 270 was an *Eco*RI site that presumably did not methylate during the preparation of the cDNA clone bank. Accordingly, no full-length cDNA clones for human CA VI were detected. To sequence the 3′-region of the molecule, the method using rapid amplification of mRNA 3′-ends (Frohman et al., 1988) was used. A synthetic oligonucleotide corresponding to a previously determined sequence near the 3′-end, the $(dT)_{17}$-adaptor ol-

```
                                                                      1
       Met Arg Ala Leu Val Leu Leu Leu Ser Leu Phe Leu Leu Gly Gly Gln Ala Gln
5'CACC ATG AGG GCC CTG GTG CTT CTG CTG TCC CTG TTC CTG CTG GGT GGC CAG GCC CAG

                                                                     21
His Val Ser Asp Trp Thr Tyr Ser Glu Gly Ala Leu Asp Glu Ala His Trp Pro Gln His
CAT GTG TCT GAC TGG ACC TAC TCA GAA GGG GCA CTG GAC GAA GCG CAC TGG CCA CAG CAC

                                                                     41
Tyr Pro Ala Cys Gly Gly Gln Arg Gln Ser Pro Ile Asn Leu Gln Arg Thr Lys Val Arg
TAC CCC GCC TGT GGG GGC CAG AGA CAG TCG CCT ATC AAC CTA CAG AGG ACG AAG GTG CGG

                                                                     61
Tyr Asn Pro Ser Leu Lys Gly Leu Asn Met Thr Gly Tyr Glu Thr Gln Ala Gly Glu Phe
TAC AAC CCC TCC TTG AAG GGG CTC AAT ATG ACA GGC TAT GAG ACC CAG GCA GGG GAG TTC

                                                                     81
Pro Met Val Asn Asn Gly His Thr Val Gln Ile Gly Leu Pro Ser Thr Met Arg Met Thr
CCC ATG GTC AAC AAT GGC CAC ACA GTG CAG ATC GGC CTG CCC TCC ACC ATG CGC ATG ACA

                                                                    101
Val Ala Asp Gly Ile Val Tyr Ile Ala Gln Gln Met His Phe His Trp Gly Gly Ala Ser
GTG GCT GAC GGC ATT GTA TAC ATA GCC CAG CAG ATG CAC TTT CAC TGG GGA GGT GCG TCC

                                                                    121
Ser Glu Ile Ser Gly Ser Glu His Thr Val Asp Gly Ile Arg His Val Ile Glu Ile His
TCG GAG ATC AGC GGC TCT GAG CAC ACC GTG GAC GGG ATC AGA CAT GTG ATC GAG ATT CAC

                                                                    141
Ile Val His Tyr Asn Ser Lys Tyr Lys Thr Tyr Asp Ile Ala Gln Asp Ala Pro Asp Gly
ATT GTT CAC TAC AAT TCT AAA TAC AAG ACG TAT GAT ATA GCC CAA GAT GCG CCG GAT GGT

                                                                    161
Leu Ala Val Leu Ala Ala Phe Val Glu Val Lys Asn Tyr Pro Glu Asn Thr Tyr Tyr Ser
TTG GCT GTA CTG GCA GCC TTC GTT GAG GTG AAG AAT TAC CCT GAA AAC ACT TAT TAC AGC

                                                                    181
Asn Phe Ile Ser His Leu Ala Asn Ile Lys Tyr Pro Gly Gln Arg Thr Thr Leu Thr Gly
AAC TTC ATT TCT CAT CTG GCC AAC ATC AAG TAC CCA GGA CAA AGA ACA ACC CTG ACT GGC

                                                                    201
Leu Asp Val Gln Asp Met Leu Pro Arg Asn Leu Gln His Tyr Tyr Thr Tyr His Gly Ser
CTT GAC GTT CAG GAC ATG CTG CCC AGG AAC CTC CAG CAC TAC TAC ACC TAC CAT GGC TCA

                                                                    221
Leu Thr Thr Pro Pro Cys Thr Glu Asn Val His Trp Phe Val Leu Ala Asp Phe Val Lys
CTC ACC ACG CCT CCC TGC ACT GAG AAC GTC CAC TGG TTT GTG CTG GCA GAT TTT GTC AAG

                                                                    241
Leu Ser Arg Thr Gln Val Trp Lys Leu Gln Asn Ser Leu Leu Asp His Arg Asn Lys Thr
CTC TCC AGG ACA CAG GTT TGG AAG CTG GAG AAT TCC TTA CTG GAT CAC CGC AAC AAG ACC

                                                                    261
Ile His Asn Asp Tyr Arg Arg Thr Gln Pro Leu Asn His Arg Val Val Glu Ser Asn Phe
ATC CAC AAC GAT TAC CGC AGG ACC CAG CCC CTG AAC CAC AGA GTG GTG GAA TCC AAC TTC

                                                                    281
Pro Asn Gln Glu Tyr Thr Leu Gly Ser Glu Phe Gln Phe Tyr Leu His Lys Ile Glu Glu
CCG AAT CAG GAA TAC ACT CTA GGC TCT GAA TTC CAG TTT TAC CTA CAT AAG ATT GAG GAA

Ile Leu Asp Tyr Leu Arg Arg Ala Leu Asn End
ATT CTT GAC TAC TTA AGA AGA GCA TTG AAC TGA GGAAAGCTAAGAGGAAGATTCAATATTAACTAGCT
TGAAGCCTGACCTAGCCAGAAGTGCCTGTCCGCTGCAGCCGCACCCTACCTTGTCTAAGAAACCATGTGTGTCTGGAAC
ACGCTGCTCCCCCTGGGGCAGCTGTTGGGATTCTGATTAAAAGAGGGGAAACGATCATCCTGGACAGGAAGTGAGATGG
CTTCAGTTCATGAGACGGGATCTGAGTTAGACATCACCAGTGGAAATTGATTGGAATAGAAACTTAAAGGAAATGGAAC
CCTAACTATTCTCCCATCAAATCATATATGTTGACCTGTCTGAATTATAAACCAGCCTGACCTTTCCTTTAGCATTAGA
TGTAATAAAATAACTTTGGAAATTTGTCATTTAAAAAAAAAAAAAAAAAAAAA.....
```

FIGURE 2: Nucleotide sequence of the human CA VI cDNA and the derived amino acid sequence for human CA VI. The putative leader sequence is underlined and the amino-terminal amino acid of the mature protein is labeled as 1. The potential glycosylation sites are indicated by the triangles—solid where it corresponds to the known glycosylation sites of sheep CA VI and hollow where not. The Cys residues, which in sheep CA VI form an intramolecular disulfide bond, are circled. The putative polyadenylation signal, AAUAAA, is boxed.

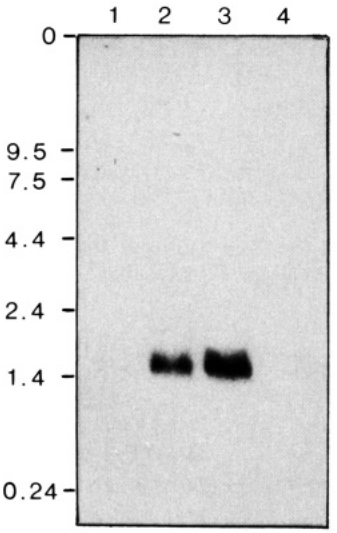

FIGURE 3: Expression of CA VI specific mRNA in different human tissues determined by Northern analysis. Total RNA from human kidney (18 μg) (lane 1), parotid gland (10 μg) (lane 2), submandibular gland (10 μg) (lane 3), and sublingual gland (10 μg) (lane 4) was analyzed as described under Materials and Methods. A ^{32}P-labeled oligonucleotide was used as a probe and the positions of the RNA standards are shown (in kilobases) on the left.

igonucleotide, and adaptor oligonucleotide were used to prime the synthesis of cDNA from the submandibular gland mRNA. This cDNA was amplified by the polymerase chain reaction, cloned, and sequenced. The strategy for determining the nucleotide sequence of the clones is summarized in Figure 1.

Nucleotide and Deduced Amino Acid Sequence. The nucleotide sequence and deduced amino acid sequence for human CA VI are shown in Figure 2. Translation is assumed to begin at the first Met residue shown, as this is immediately preceded by part of the translation initiation consensus sequence—CACC (Kozak, 1987). This is followed by a leader sequence typical of secreted proteins (Austen, 1979) with an Arg at the amino-terminal end and then a region rich in hydrophobic residues. A polyadenylation consensus signal occurs 29 bases upstream from the poly(A+) tail (Proudfoot & Brownlee, 1976), giving a total of 1314 bases in the sequence. This, with the poly(A+) tail and 5′-untranslated sequence not included in these clones, agrees well with the size of 1.45 kb determined by Northern analysis (Figure 3) for this mRNA species.

The protein isolated from human saliva has a blocked amino terminus (Fernley et al., 1989). The residue in human CA VI corresponding to the amino-terminal glycine in sheep CA VI is a glutamine. Presumably, the leader peptidase cleaves the preprotein after an alanine residue [as would be predicted (Austen, 1979)] leaving an amino-terminal glutamine residue. This would undergo cyclization to form a pyroglutamate residue that is resistant to Edman degradation. The mature protein would then have 291 amino acids with a calculated M_r of 33 646 and this is close to the measured M_r (36 000) of the deglycosylated protein (Murakami & Sly, 1987). Overall, human CA VI has a 71.5% sequence identity with sheep CA VI (Figure 4). The amino acid sequences of four of the isozymes of human carbonic anhydrase are compared in Figure 5. Only 63 residues (24%) are common to all four isozymes, although many other positions have conservative substitutions. CA VI has a 35% sequence identity with CA II, which is the high-activity isozyme.

Human CA VI is a glycoprotein and has two N-linked oligosaccharide chains (Murakami & Sly, 1987). The amino acid sequence reveals three potential N-glycosylation sites (Asn-X-Thr/Ser) and two of these, Asn 50 and Asn 239, are known to be glycosylated in sheep CA VI (Fernley et al., 1988b). Sheep CA VI has an intramolecular disulfide bond linking Cys 25 and Cys 207. These two cysteines are conserved in the human enzyme and they presumably form a disulfide bond here also.

Northern Analysis and Hybridization Histochemistry. With a partial cDNA clone as a probe, parotid and submandibular gland mRNA showed bands at 1.45 kb, whereas the

```
HCAVI  Q H V S D W T Y S E G A L D E A H W P Q H Y P A C G G Q R Q S P I N L Q R T K
OCAVI  G H G V E W T Y S E G M L D E A H W P L E Y P K C G G R R Q S P I D L Q M K K

       V R Y N P S L K G L N M T G Y E T Q A G E F P M V N N G H T V Q I G L P S T M
       V Q Y N P S L R A L N L T G Y G L W H G E F P M T N N G H T V Q I S L P S T M

       R M T V A D G I V Y I A Q Q M H F H W G G A S S E I S G S E H T V D G I R H V
       S M T T S D G T Q Y L A K Q M H F H W G G A S S E I S G S E H T V D G M R Y V

       I E I H I V H Y N S K Y K T Y D I A Q D A P D G L A V L A A F V E V K N Y P E
       I E I H V V H Y N S K Y N S Y E E A Q K E P D G L A V L A A L V E V K D Y T E

       N T Y Y S N F I S H L A N I K Y P G Q R T T L T G L D V Q D M L P R N L Q H Y
       N A Y Y S K F I S H L E D I R Y A G Q S T V L R G L D I E D M L P G D L R Y Y

       Y T Y H G S L T T P P C T E N V H W F V L A D F V K L S R T Q V W K L E N S L
       Y S Y L G S L T T P P C T E N V H W F V V A D T V K L S K T Q V E K L E N S L

       L D H R N K T I H N D Y R R T Q P L N H R V V E S N F - - - P N Q E Y T L G S
       L N H Q N K T I Q N D Y R R T Q P L N H R V V E A N F M S R P H Q E Y T L A S

       E F Q F Y L H K I E E I L D Y L R R A L N
       K L H F Y L N N I D Q T L E Y L R R F I E Q K I P K R K K Q E N W P
```

FIGURE 4: Comparison of the amino acid sequences of human and sheep carbonic anhydrase VI. Residues that are conserved are boxed.

```
                1                                                  20                                                  40
H CAI    :      A S P D W G Y D D K N G P E - Q W S K L Y P I A N G N N Q S P V D I K T S E T K H D T S L K
H CAII   :        S H H W G Y G K H N G P E - H W H K D F P I A K G E R Q S P V D I D T H T A K Y D P S L K
H CAIII  :        A K E W G Y A S H N G P D - H W H E L F P N A K G E N Q S P I E L H T K D I R H D P S L Q
H CAVI   :   Q H V S D W T Y S E G A L D E A H W P Q H Y P A C G G Q R Q S P I N L Q R T K V R Y N P S L K

                                                   60                                                  80
I        :   P I S V S - Y N P A T A K E - I I N V G H S F H V N F E D N D N R S V L K G G P F S D S Y R L
II       :   P L S V S - Y D Q A T S L R - I L N N G H A F N V E F D D S Q D K A V L K G G P L D G T Y R L
III      :   P W S V S - Y D G G S A K T - I L N N G K T C R V V F D D T Y D R S M L R G G P L P G P Y R L
VI       :   G L N M T G Y E T Q A G E F P M V N N G H T V Q I G L P S T M R M T V A - D G - - - I V Y I A

                                          100                                                 120
I        :   F Q F H F H W G - - S T N E H G S E H T V D G V K Y S A E L H V A H W N S A K Y S S L A E A A
II       :   I Q F H F H W G - - S L D G Q G S E H T V D K K K Y A A E L H L V H W N T - K Y G D F G K A V
III      :   R Q F H L H W G - - S S D D H G S E H T V D G V K Y A A E L H L V H W N P - K Y N T F K E A L
VI       :   Q Q M H F H W G G A S S E I S G S E H T V D G I R H V I E I H I V H Y N S - K Y K T Y D I A Q

                    140                                                 160                                                 180
I        :   S K A D G L A V I G V L M - - K V G E A N P K L Q K V L D A L Q A I K T K G K R A P F T N F D
II       :   Q Q P D G L A V L G I F L - - K V G S A K P G L Q K V V D V L D S I K T K G K S A D F T N F D
III      :   K Q R D G I A V I G I F L - - K I G H E N G E F Q I F L D A L D K I K T K G K E A P F T K F D
VI       :   D A P D G L A V L A A F V E V K N Y P E N T Y Y S N F I S H L A N I K Y P G Q R T T L T G L D

                                                   200                                                 220
I        :   P S T L L P S S L - D F W T Y P G S L T H P P L Y E S V T W I I C K E S I S V S S E Q L A Q F
II       :   P R G L L P E S L - D Y W T Y P G S L T T P P L L E C V T W I V L K E P I S V S S E Q V L K F
III      :   P S C L F P A C R - D Y W T Y Q G S F T T P P C E E C I V W L L L K E P M T V S S D Q M A K L
VI       :   V Q D M L P R N L Q H Y Y T Y H G S L T T P P C T E N V H W F V L A D F V K L S R T Q V W K L

                                          240                                                 260
I        :   - R S L L S N V E G D N A V P M Q H N N R P T Q P L K G R T V R A S F
II       :   - R K L N F N G E G E P E E L M V D N W R P A Q P L K N R Q I K A S F K
III      :   - R S L L S S A E N E P P V P L V S N W R P P Q P I N N R V V R A S F K
VI       :   E N S L L D H R N K T I H - - - - N D Y R R T Q P L N H R V V E S N F P N Q E Y T L G S E F Q

VI       :   F Y L H K I E E I L D Y L R R A L N
```

FIGURE 5: Comparison of the amino acid sequences of human carbonic anhydrases I, II, III, and VI. Residues that are conserved in all four isozymes are boxed. Residues postulated to be involved in binding to the active-site Zn, to the Zn–ligand His residues, or to the Zn-bound solvent molecules are marked by an asterisk. Gaps have been introduced into the sequences to optimize homology. The numbering system used here is based on that of human CA I so comparisons of active-site residues can be made. The CA I sequence is from Barlow et al. (1987), the CA II sequence is from Montgomery et al. (1987), and the CA III sequence from Wade et al. (1986). This latter sequence is identical with that reported by Lloyd et al. (1986) except that there is an allelic variation at position at 31 (valine instead of isoleucine).

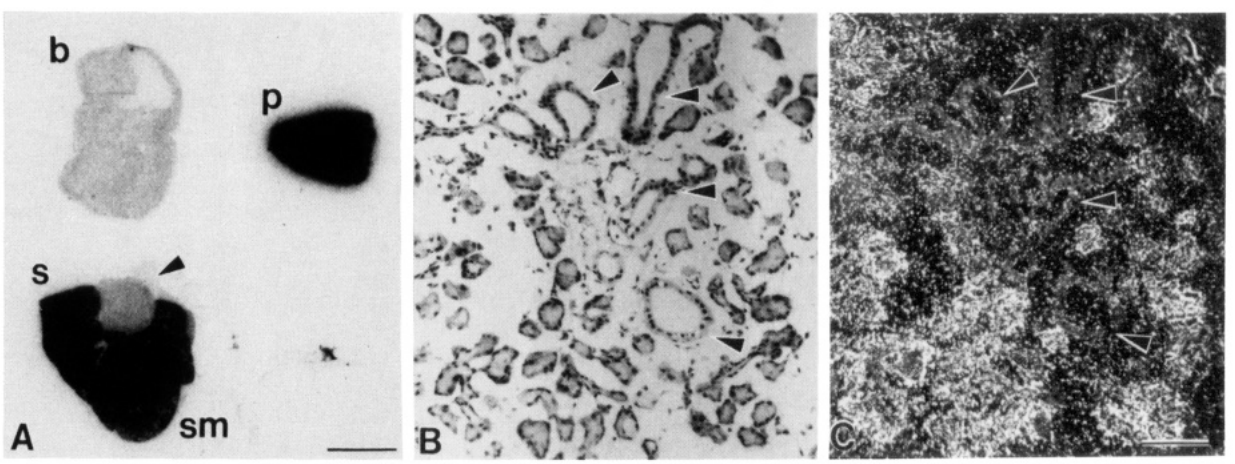

FIGURE 6: Autoradiographs on X-ray film (A) or liquid emulsion (B, C) of 6-μm frozen sections after hybridization with a ^{32}P-labeled oligonuclotide 30-mer complementary to human CA VI mRNA. (A) Mouse brain (b), human parotid gland (p), and mouse salivary glands (s) with the submandibular gland marked (sm) and the sublingual gland arrowed. Bar, 3 mm. (B, C) The same field of the human parotid gland photographed in bright field (B) and dark field (C). Silver grains in the emulsion overlaying the acinar tissue can be seen as white dots in (C). Ducts (arrowed) are negative. Stain: haematoxylin and eosin. Bar, 0.5 mm.

sublingual gland and the kidney mRNAs showed no hybridization (Figure 3). Prostate and liver mRNA also showed no hybridization with this probe (results not shown). With hybridization histochemistry, the human parotid and mouse submandibular glands were labeled strongly while the sublingual mouse salivary gland and brain sections did not hybridize (Figure 6).

Southern Analysis. Southern blot analysis of human genomic DNA digested individually with *Bam*HI, *Sac*I, *Bgl*II, *Hin*dIII, and *Eco*RI exhibited single hybridization-positive fragments (Figure 7). No additional hybridization-positive bands were detected, suggesting there are no other closely related genes or pseudogenes in the human clone. Thus human CA VI appears to be encoded by a single-copy gene.

Discussion

This is the first reported nucleotide sequence for a secreted carbonic anhydrase. It was first suggested by Feldstein and Silverman (1984) that this enzyme was secreted by the salivary glands and the view is confirmed by the demonstration of a leader sequence typical of secreted proteins. Apart from this isozyme and the cytoplasmic isozymes, distinct membrane-bound and mitochondrial carbonic anhydrases are known to occur (Wistrand, 1984; Storey et al., 1984). Thus the carbonic

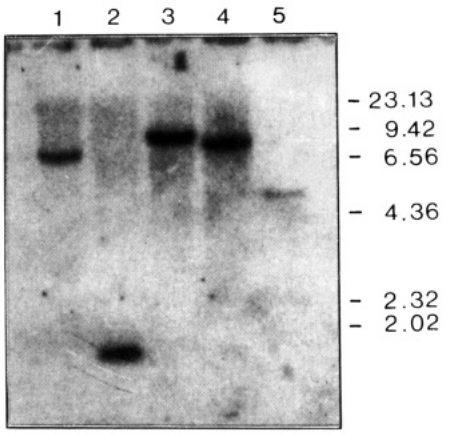

FIGURE 7: Southern blot analysis of human genomic DNA (10 μg each) digested with the following restriction endonucleases, *Bam*HI (lane 1), *Sac*I (lane 2), *Bgl*II (lane 3), *Hin*dIII (lane 4), and *Eco*RI (5 μg, lane 5), electrophoresed on an 0.8% agarose gel, transferred to a nylon membrane, and probed as described under Materials and Methods. Size markers are a *Hin*dIII digest of phage λ DNA (length in kilobases).

anhydrases form a large gene family whose transcription products have a wide subcellular distribution. While the cytoplasmic isozymes are related closely to one another (typically with about 60% identity), the salivary CAs have only about 34% sequence identity with (e.g.) CA II. This suggests that the salivary isozyme gene diverged from the ancestral CA gene at a relatively early stage of evolution. The human and sheep CA VI have a high degree of sequence identity (71.5%) and studies with CA VI antibodies (Fernley et al., 1989) suggest that many mammalian CA VIs are related closely. The structural similarity of the two secreted CA isozymes is also reflected in their hydrophobicity plots (results not known). These are very similar except for the carboxy-terminal region. The sheep enzyme has 16 extra amino acids in this region and is very hydrophilic. The function of this region is not known.

Sixty-three residues are conserved in all so-far sequenced isozymes of human CA (Figure 5). If all other sequenced CAs are included in this analysis, the number of "invariant" residues falls to 51. As might be expected, many of these are involved in the active site of the enzyme. Of the hydrophilic residues that form part of the active site of CA (Notstrand et al., 1975), residues 7, 61, 92, 94, 96, 106, 107, 119, and 199 appear to be invariant. These include the zinc-binding histidine residues 94, 96, and 119. Residue 64 is a histidine, except in CA III where it is a lysine. However, this does not explain the low activity of CA III, as Forsman et al. (1988) have used site-directed mutagenesis to replace His 64 of CA II with Lys (or Gln, Glu, Ala) and found a fall in activity of only 1.5–3.5-fold—not the 67-fold fall expected for a CA III like enzyme. Residue 67 could be more important for high activity with Asn in CA II and Gln in CA VI (both high-activity isozymes), while the lower activity CA I has His and the very low activity CA III Arg in this position. Residue 200 is Thr in CA II, III, and VI but His in CA I, and this may contribute to the lower activity of CA I (but obviously not for CA III). The hydrophobic residues Leu 198 and Val 207 are important for substrate binding (Eriksson, 1988) and they have been conserved in CA VI but have changed to Phe and Ile, respectively, in CA III, probably accounting for much of the lowered activity of CA III.

CA VI has many small deletions and insertions in the sequence compared to the three cytoplasmic enzymes (Figure 5). These appear to be situated on the surface of the molecule and could probably be accommodated in the model for cytoplasmic carbonic anhydrase without drastic conformational changes (Eriksson, 1988). The modeling studies of Eriksson also suggest that the carbohydrate attachment sites (Asn 50 and Asn 239 in sheep CA VI and presumably human CA VI) would occur at the surface of the molecule and that the two cysteine residues would be sufficiently close to form a disulfide bond. Although human CA VI has three potential glycosylation sites, Murakami and Sly (1987) have found that only two are glycosylated.

Northern analysis confirms the conclusion reached by Western analysis (Fernley et al., 1989) that the salivary glands are the major (and possibly the only) sites for synthesis of the secreted carbonic anhydrases, although low levels of synthesis in other tissues cannot be ruled out by these methods. With the human salivary glands, both the parotid and submandibular glands have high concentrations of CA VI mRNA, whereas in the sheep there are higher concentrations of CA VI mRNA in the parotid than in the submandibular gland (Fernley et al., 1989).

For the Southern analysis, a region of the human cDNA sequence that had little sequence identity with the other carbonic anhydrase isozymes was chosen to design the oligonucleotide probes. Also, in other CA genes so far examined, the region used did not possess introns (Venta et al., 1985), which might contain restriction sites for the restriction enzymes used. The finding of one CA VI gene agrees with our earlier finding that the CA VI gene is located in one specific region on the short arm of chromosome 1 (Sutherland et al., 1989).

ACKNOWLEDGMENTS

We thank J. Close and Dr. R. Richards who made available the human submandibular gland cDNA library and Dr. J. Savige for organizing the collection of human tissue.

Registry No. CA, 9001-03-0; DNA (human carbonic anhydrase isoenzyme VI of mRNA complementary), 130405-56-0; carbonic anhydrase (human isoenzyme VI precursor protein moiety reduced), 130405-61-7; carbonic anhydrase (human isoenzyme VI precursor protein moiety reduced pGlu-1), 130405-62-8.

REFERENCES

Amor-Gueret, M., & Levi-Strauss, M. (1990) *Nucleic Acids Res. 18*, 1646.

Austen, B. M. (1979) *FEBS Lett. 103*, 308–313.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Barlow, J. H., Lowe, N., Edwards, Y. H., & Butterworth, P. H. W. (1987) *Nucleic Acids Res. 15*, 2386.

Benton, W. D., & Davis, R. W. (1977) *Science 196*, 180–182.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Clouston, W. M., Fournier, R. E. K., & Richards, R. I. (1989) *FEBS Lett. 255*, 419–422.

Edwards, Y. (1990) *Biochem. Soc. Trans. U.K. 18*, 171–175.

Eriksson, A. E. (1988) *Acta Univ. Upsaliensis 164* (summary of Ph.D. Thesis).

Evans, B. A., Yun, Z. X., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., Callen, D. F., Baker, E., Hyland, V. J., Sutherland, G. R., & Richards, R. I. (1988) *Biochemistry 27*, 3124–3129.

Feldstein, J. B., & Silverman, D. N. (1984) *J. Biol. Chem. 259*, 5447–5453.

Fernley, R. T. (1988) *Trends Biochem. 13*, 356–359.

Fernley, R. T., Wright, R. D., & Coghlan, J. P. (1979) *FEBS Lett. 105*, 299–302.

Fernley, R. T., Congiu, M., Wright, R. D., & Coghlan, J. P. (1984) *Ann. N.Y. Acad. Sci. 429*, 212–213.

Fernley, R. T., Coghlan, J. P., & Wright, R. D. (1988a) *Biochem. J. 249*, 201-207.
Fernley, R. T., Wright, R. D., & Coghlan, J. P. (1988b) *Biochemistry 27*, 2815-2820.
Fernley, R. T., Darling, P., Aldred, P., Wright, R. D., & Coghlan, J. P. (1989) *Biochem. J. 259*, 91-96.
Forsman, C., Behravan, G., Jonsson, B.-H., Liang, Z.-W., Lindskog, S., Ren, X., Sandstrom, J., & Wallgren, K. (1988) *FEBS Lett. 229*, 360-362.
Frohman, M. A., Dush, M. K., & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8998-9002.
Kozak, M. (1987) *Nucleic Acids Res. 15*, 8125-8148.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105-132.
Lloyd, J., McMillan, S., Hopkinson, D., & Edwards, Y. (1986) *Gene 41*, 233-239.
Maniatis, T., Fritsch, E. F., & Sambrook, J., (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Montgomery, J. C., Venta, P. J., Tashian, R. E., & Hewett-Emmett, D. (1987) *Nucleic Acids Res. 15*, 4687.
Murakami, H., & Sly, W. S. (1987) *J. Biol. Chem. 262*, 1382-1388.
Notstrand, B., Vaara, I., & Kannan, K. K. (1975) in *Isozymes: Molecular Structure* (Markert, C. L., Ed.) Vol. 1, pp 575-599, Academic Press, New York.
Penschow, J. D., Haralambidis, J., Aldred, P., Tregear, G. W., & Coghlan, J. P. (1986) *Methods Enzymol. 124*, 534-548.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211-214.
Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487-491.
Sanger, F., Nicklin, S., & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503-517.
Staden, R. (1977) *Nucleic Acids Res. 10*, 4731-4751.
Storey, B. T., Dodgson, S. J., & Forster, R. E., II (1984) *Ann. N.Y. Acad. Sci. 429*, 210-211.
Sutherland, G. R., Baker, E., Fernandez, K. E. W., Callen, D. F., Aldred, P., Coghlan, J. P., Wright, R. D., & Fernley, R. T. (1989) *Cytogenet. Cell Genet. 50*, 149-150.
Tashian, R. E. (1989) *BioEssays 10*, 186-192.
Venta, P. J., Montgomery, J. C., Hewett-Emmett, D., Wiebauer, K., & Tashian, R. E. (1985) *J. Biol. Chem. 260*, 12130-12135.
Venta, P. J., Montgomery, J. C., & Tashian, R. E. (1987) *Isozymes: Curr. Top. Biol. Med. Res. 14*, 59-72.
Wade, R., Gunning, P., Eddy, R., Shows, T., & Kedes. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9571-9575.
Wistrand, P. J. (1984) *Ann. N.Y. Acad. Sci. 429*, 195-206.

Solution Structure of Human Calcitonin Gene-Related Peptide by ^{1}H NMR and Distance Geometry with Restrained Molecular Dynamics†

Alexander L. Breeze, Timothy S. Harvey, Renzo Bazzo, and Iain D. Campbell*
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.
Received June 5, 1990; Revised Manuscript Received September 11, 1990

ABSTRACT: The structure of human calcitonin gene-related peptide 1 (hCGRP-1) has been determined by ^{1}H NMR in a mixed-solvent system of 50% trifluoroethanol/50% H_2O at pH 3.7 and 27 °C. Complete resonance assignment was achieved by using two-dimensional methods. Distance restraints for structure calculations were obtained by semiquantitative analysis of intra- and interresidue nuclear Overhauser effects; in addition, stereospecific or χ^1 rotamer assignments were obtained for certain side chains. Structures were generated from the distance restraints by distance geometry, followed by refinement using molecular dynamics, and were compared with experimental NH–CαH coupling constants and amide hydrogen exchange data. The structure of hCGRP-1 in this solvent comprises an amino-terminal disulfide-bonded loop (residues 2–7) leading into a well-defined α-helix between residues 8 and 18; thereafter, the structure is predominantly disordered, although there are indications of a preference for a turn-type conformation between residues 19 and 21. Comparison of spectra for the homologous hCGRP-2 with those of hCGRP-1 indicates that the conformations of these two forms are essentially identical.

Calcitonin gene-related peptide (CGRP)[1] is a 37 amino acid residue single-chain polypeptide which is characterized by a carboxy-terminal phenylalanine amide and an amino-terminal disulfide-bonded loop. Two forms of CGRP (designated α- and β-CGRP, or CGRP-1 and CGRP-2) were predicted to arise as alternative splicing products of the α- and β-calcitonin gene complexes (Amara et al., 1982; Rosenfeld et al., 1983; Hoppener et al., 1985; Steenbergh et al., 1985); the existence of CGRP-1 in man was confirmed following its isolation and characterization by fast atom bombardment mass spectrometry (Morris et al., 1984). The peptide was subsequently found to have a widespread distribution in the body, being particularly prevalent in the central nervous system, and peripherally in nerves of the cardiovascular system (Gibson et al., 1984).

† This is a contribution from the Oxford Centre for Molecular Sciences which is supported by the SERC and MRC. A.L.B. and T.S.H. were supported additionally by ICI.
* Address correspondence to this author.

[1] Abbreviations: CGRP, calcitonin gene-related peptide; CD, circular dichroism; NMR, nuclear magnetic resonance; hCGRP, human CGRP; DG, distance geometry; REM, restrained energy minimization; RMD, restrained molecular dynamics; TFE, trifluoroethanol; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; DQF-COSY, double-quantum-filtered two-dimensional correlation spectroscopy; E.COSY, exclusive correlation spectroscopy; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; TPPI, time-proportional phase incrementation; rms, root mean square.

0006-2960/91/0430-575$02.50/0